

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-268893

(43)Date of publication of application : 09.10.1998

(51)Int.Cl. G10L 3/00  
G10L 3/00

(21)Application number : 09-091607

(71)Applicant : NEC CORP

(22)Date of filing : 26.03.1997

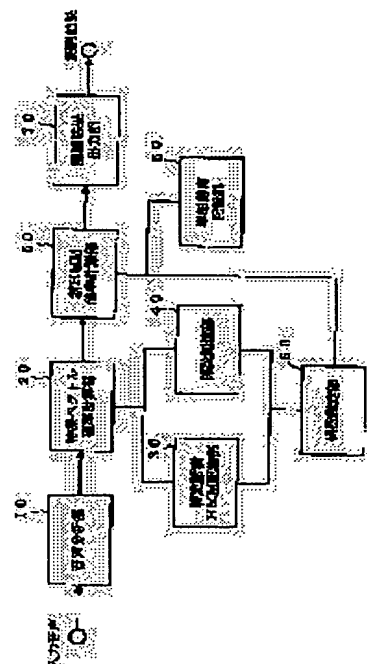
(72)Inventor : ISO KENICHI

## (54) VOICE RECOGNITION DEVICE

### (57)Abstract:

**PROBLEM TO BE SOLVED:** To prepare a new HMM optimum to a new speaker, by making the linear combination of the output probability of the HMM read to from an HMM storage part according to an input voice the output probability of the new HMM, and deciding the coefficient of the linear combination so as to maximize the probability of the new HMM.

**SOLUTION:** A specified speaker HMM storage part 30 stores respective specified speaker HMMs of plural speakers. The specified speaker HMMs of respective speakers are shown by the output probability and a transit probability. A coefficient storage part 40 stores a coefficient for linearly connecting the HMMs of respective speakers. A characteristic vector output probability calculation part 20 calculates the output probability of all states of all speakers for input voice characteristic vectors of respective times, and calculates the output probability for all states of the new HMM as the linear combination. A new HMM probability calculation part 60 calculates the probability of the new HMM for the input voice. A coefficient estimation part 80 estimates the coefficient so as to maximize the probability for the input voice calculated by the new HMM probability calculation part 60.



## LEGAL STATUS

[Date of request for examination] 23.04.1997

[Date of sending the examiner's decision of rejection] 06.06.2000

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3144341

[Date of registration] 05.01.2001

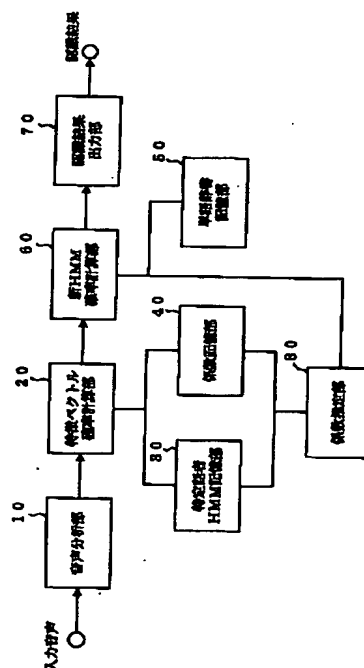
[Number of appeal against examiner's decision of rejection] 2000-10292

[Date of requesting appeal against examiner's decision of rejection] 06.07.2000

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(11)特許出願公開番号



## 【特許請求の範囲】

【請求項1】HMMを用いた音声認識装置において、複数のHMMを格納したHMM記憶部と、入力音声に対応して前記HMM記憶部から読み出した複数のHMMの出力確率の線形結合を新HMMの出力確率とし、前記新HMMの確率を最大化あるいは極大化するように前記線形結合の係数を決定する係数推定部とを備えて成ることを特徴とする音声認識装置。

【請求項2】前記係数推定部は、前記入力音声の発話内容を既知として、対応する新HMMの確率を最大化あるいは極大化するように前記線形結合の係数を決定する請求項1に記載の音声認識装置。

【請求項3】発話内容未知の入力音声に対して、認識対象単語辞書の各単語の新HMM確率を最大化あるいは極大化するように前記線形結合の係数を推定する係数推定部と、最大の確率を与えた単語を認識結果として出力する認識結果出力部とを有する請求項1に記載の音声認識装置。

【請求項4】新話者のHMMの出力確率と遷移確率として、予め用意した多数の話者の特定話者HMMの各出力確率及び遷移確率を、新話者の各特定話者に対する類似度パラメータで線形結合した出力確率と遷移確率を用いて入力音声の認識を行う音声認識装置において、前記新話者の各特定話者に対する類似度パラメータを、新話者の未知あるいは既知の発声に対する新話者HMMの尤度が最大または極大になるように最適に推定することを特徴とする音声認識装置。

【請求項5】前記最適推定は、前記複数のHMMの出力確率の線形結合を新HMMの出力確率とし、前記入力音声に対する新HMMの確率を最大化あるいは極大化するように前記線形結合の係数を決定する請求項4に記載の音声認識装置。

【請求項6】入力音声进行分析し、一定の時間間隔ごと音響特徴ベクトルを求める音声分析部と、複数の話者のそれぞれの特定話者HMMを記憶する特定話者HMM記憶部と、各話者のHMMを線形結合するための係数を記憶する係数記憶部と、各時刻の入力音声特徴ベクトル $t$ に対する、全話者の全状態の出力確率を算出して、その線形結合として新HMMの全状態に対する出力確率を算出する特徴ベクトル出力確率計算部と、認識対象単語のそれぞれに対して、各単語がどのようなHMMの状態列で表わされるかを記憶する単語辞書記憶部と、前記入力音声に対する新HMMの確率を算出する新HMM確率計算部と、前記新HMM確率計算部からの新HMM確率の最大値を与える単語を認識結果として出力する認識結果出力部と、

前記新HMM確率計算部で算出された入力音声に対する確率を最大化あるいは極大化するような係数を推定し、推定された係数を前記係数記憶部に記憶する係数推定部と、を備えて成ることを特徴とする音声認識装置。

## 【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、音声認識装置に関し、特に不特定話者に対する認識率を改善した音声認識装置に関するものである。

10 【0002】

【従来の技術】新しい話者に対して少ない発声で音声認識装置を話者適応化する方式として、予め複数（多数）の話者のそれぞれに対する特定話者標準パターン（HMM、隠れマルコフモデル）を用意しておき、その中から上記新話者の発声に類似した特定話者標準パターンを1個または複数個選択して用いる方式（特開平1-161399号公報）や、新話者と予め用意した複数の話者のそれぞれとの類似度を算定する手段を用意しておき、新話者の標準パターンとして、各話者への類似度を荷重係数として各特定話者標準パターンを重ね合わせた標準パターンを生成する方式（特開平4-121793号公報）が知られている。

【0003】

【発明が解決しようとする課題】前者の方式（特開平1-161399号公報）においては、予め用意した複数の特定話者の中に、新話者に類似した話者が含まれていなかった場合は、無理失理にあまり類似していない話者が選ばれてしまうことがあり、精度の高い話者適応化は望めない。

30 【0004】また、後者の方式（特開平4-121793号公報）においては、類似度として予め各特定話者の任意の発声を用いて、各話者ごとに音響特徴ベクトルの連鎖に対する確率分布 $P(X_t | X_{t-1}, X_{t-2}, X_{t-3}, \dots)$ （例えば、音響特徴ベクトルをベクトル量子化した場合はコード番号列に対する確率テーブルになる）を用意しておき、新話者の発声の音響特徴ベクトル列に対する確率を、各話者の確率分布を用いて算出した値をその話者への類似度としている。この方式は、新話者の発話内容が未知の場合も適用可能であるという利点があるが、発話内容が既知の場合もその情報を活用できず、また発話内容の既知・未知に関わらず、必ずしも数学的に最適な類似度算定方法という保証がないために高い精度は望めない。

【0005】

【課題を解決するための手段】前述の課題を解決するため、本発明による音声認識装置は、HMMを用いた音声認識装置において、複数のHMMを格納したHMM記憶部と、入力音声に対応して前記HMM記憶部から読み出した複数のHMMの出力確率の線形結合を新HMMの出力確率とし、前記新HMMの確率を最大化あるいは極大

化するように前記線形結合の係数を決定する係数推定部とを備えて構成される。

【0006】ここで、前記係数推定部は、前記入力音声の発話内容を既知として、対応する新HMMの確率を最大化あるいは極大化するように前記線形結合の係数を決定し、また、発話内容未知の入力音声に対して、認識対象単語辞書の各単語の新HMM確率を最大化あるいは極大化するように前記線形結合の係数を推定する係数推定部と、最大の確率を与えた単語を認識結果として出力する認識結果出力部とを有するように構成される。

【0007】また、本発明の音声認識装置では、新話者のHMMの出力確率と遷移確率として、予め用意した多数の話者の特定話者HMMの各出力確率及び遷移確率を、新話者の各特定話者に対する類似度パラメータで線形結合した出力確率と遷移確率を用いて入力音声の認識を行う音声認識装置において、前記新話者の各特定話者に対する類似度パラメータを、新話者の未知あるいは既知の発声に対する新話者HMMの尤度が最大または極大になるように最適に推定するように構成される。

【0008】ここで、前記最適推定は、前記複数のHMMの出力確率の線形結合を新HMMの出力確率とし、前記入力音声に対する新HMMの確率を最大化あるいは極大化するように前記線形結合の係数を決定する。

【0009】更に、本発明の他の態様による音声認識装置は、入力音声进行分析し、一定の時間間隔ごと音響特徴ベクトルを求める音声分析部と、複数の話者のそれぞれの特定話者HMMを記憶する特定話者HMM記憶部と、各話者のHMMを線形結合するための係数を記憶する係数記憶部と、各時刻の入力音声特徴ベクトル $x_t$ に対する、全話者の全状態の出力確率を算出して、その線形結合として新HMMの全状態に対する出力確率を算出する特徴ベクトル出力確率計算部と、認識対象単語のそれぞれに対して、各単語がどのようなHMMの状態列で表わされるかを記憶する単語辞書記憶部と、前記入力音声に対する新HMMの確率を算出する新HMM確率計算部と、前記新HMM確率計算部からの新HMM確率の最大値を与える単語を認識結果として出力する認識結果出力部と、前記新HMM確率計算部で算出された入力音声に対する確率を最大化あるいは極大化するような係数を推定し、推定された係数を前記係数記憶部に記憶する係数推定部とを備えて構成される。

【0010】

【発明の実施の形態】図1は本発明に基づく音声認識装置の一実施形態の構成ブロック図である。音声分析部10は、入力音声进行分析し、一定の時間間隔ごと（例えば、10ミリ秒ごと）に抽出したケプストラムなどの音響特徴ベクトル $X_t$ を求め、特徴ベクトル出力確率計算部20に送出する。ここで、音響特徴ベクトル $X_t$ は、例えば、10次元のケプストラム・ベクトルで、添字 $t$ は時間順序を表わす番号（自然数）である。一回の入力

発声に対する音響特徴ベクトル時系列全体を $X$ で表わす。

$$X = x_1, x_2, \dots, x_t, \dots, x_T$$

【0011】特定話者HMM記憶部30には、複数の話者のそれぞれの特定話者HMMを記憶する。HMMは隠れマルコフモデルの意味で、音声認識分野で最も一般的に知られ、使用されている認識方式（モデル）であり、詳細は文献「音声認識の基礎（上・下）、古井 監訳、NTTアドバンステクノロジー株式会社」（原本は英語で“Fundamentals of Speech Recognition”, L. Rabiner and B-H Juang, Prentice Hall）に詳しい。ここでは、例えば、特定話者のHMMは、その話者の大量の音声データから学習によって構築したもので、音素HMMであるとする（音素とは単語より小さい音声の単位で、単語や文のHMMは音素HMMの連結で表わされる）。

【0012】各話者の特定話者HMMは、出力確率 $b_{i, (x)}$ と遷移確率 $a_{i, i'}$ で表わされる。ここで、添字 $s$ は話者を表わす番号（自然数）で全話者数を $S$ 人すると、 $s = 1, 2, \dots, S$ となる。添字 $i$ と添字 $j$ は、HMMの状態を表わす番号（自然数）で全状態数を $N$ 個とすると、 $i, j = 1, \dots, N$ となる。出力確率 $b_{i, (x)}$ は、話者 $s$ の特定話者HMMの状態 $i$ が音響特徴ベクトル $x$ を出力する確率を表わす。遷移確率 $a_{i, i'}$ は、話者 $s$ の特定話者HMMの状態 $i$ から状態 $j$ への遷移確率である。これらを用いることにより、入力音声 $X$ に対する話者 $s$ の特定話者HMMの確率 $P(X, s)$ を算出することができる。

【数1】

$$P(X, s) = \sum_{q_1=1}^N \sum_{q_2=1}^N \dots \sum_{q_T=1}^N P(X, s, q_1 q_2 \dots q_t \dots q_T)$$

$$P(X, s, q_1 q_2 \dots q_t \dots q_T) = \prod_{t=1}^T a_{q_{t-1} q_t}^{(s)} b_{q_t}^{(s)}(x_t)$$

ここで、 $q_1, q_2, \dots, q_T$ は、各時刻におけるHMMの状態を表わしている。

【0013】係数記憶部40は、各話者のHMMを線形結合するための係数

$$\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_s, \dots, \lambda_S\}$$

を記憶する。 $\lambda_s$ は話者 $s$ の特定話者HMMに対する係数で、全話者に対する係数の総和は以下のように規格化されている。

【数2】

$$\sum_{s=1}^S \lambda_s = 1$$

【0014】特徴ベクトル出力確率計算部20は、各時刻の入力音声特徴ベクトル $x_t$ に対する、全話者（ $s =$

1, 2, ..., S) の全状態 ( $i = 1, \dots, N$ ) の出力確率  $b_i^{(s)}(x_t)$  を算出して、その線形結合として新HMMの全状態 ( $i = 1, \dots, N$ ) に対する出力確率【数3】

$$b_i(x_t) = \sum_{s=1}^S \lambda_s b_i^{(s)}(x_t)$$

を算出する。

【0015】単語辞書記憶部50は、認識対象単語のそれぞれに対して、各単語がどのようなHMMの状態列で表わされるかを記憶している。例えば、特定話者HMM\*

$$P_{new}(X|w, \Lambda) = \sum_{q_1=1}^N \sum_{q_2=1}^N \dots \sum_{q_T=1}^N P_{new}(X, q_1 q_2 \dots q_i \dots q_T | w, \Lambda)$$

【数6】

$$P_{new}(X, q_1 q_2 \dots q_i \dots q_T | w, \Lambda) = \prod_{t=1}^T a_{q_{t-1} q_t} b_{q_t}(x_t)$$

ここで、 $a_{q_{t-1} q_t}$  は新HMMの状態  $q_{t-1}$  から状態  $q_t$  への遷移確率  $b_{q_t}(x_t)$  は新HMMの状態  $q_t$  における音響特徴ベクトル  $x_t$  の出力確率である。この出力確率は前記の特徴ベクトル出力確率計算部20により特定話者HMMの出力確率を線形結合して算出されたものである。また、記号  $w$  は、認識対象単語の番号（自然数）を表わしている。入力音声の発話内容が既知の場合は、その発話に対応する単語のHMM状態列情報を単語辞書記憶部50から読み出して、その状態列のみが状態遷移に現れるように、遷移確率  $a_{i,1}$  のそれ以外の成分を0にする。

【0017】入力音声の発話内容が未知の場合は、単語辞書記憶部50に記憶されている全ての認識対象単語のそれぞれについて、上記の発話内容が既知の場合と同様の確率計算を行い、全ての単語に対する確率を認識結果出力部70へ送り、認識結果出力部70は、その中の最大値を与える単語

【数7】

$$P_{new}(X|w, \Lambda) = \sum_{q_1=1}^N \sum_{q_2=1}^N \dots \sum_{q_T=1}^N \sum_{s_1=1}^S \sum_{s_2=1}^S \dots \sum_{s_T=1}^S P_{new}(X, q_1 q_2 \dots q_i \dots q_T, s_1 s_2 \dots s_i \dots s_T | w, \Lambda)$$

【数11】

$$P_{new}(X, q_1 q_2 \dots q_i \dots q_T, s_1 s_2 \dots s_i \dots s_T | w, \Lambda) = \prod_{t=1}^T a_{q_{t-1} q_t} \lambda_s b_{q_t}^{(s)}(x_t)$$

ここで、次のようなQ関数を定義する。

★ ★【数12】

$$Q(\Lambda, \Lambda') = \sum_Q \sum_S P_{new}(X, Q, S | w, \Lambda) \ln [P_{new}(X, Q, S | w, \Lambda')] - \nu \left( \sum_{s=1}^S \lambda_s - 1 \right)$$

上式では表記を簡単にするために、以下の簡略表記を用いた。

【数13】

\* 記憶部30が各話者の音素HMMを記憶している場合は、各単語の音素表記を記憶している（音節HMMを用いる場合は各単語の音節表記を記憶している）。

【0016】新HMM確率計算部60は、入力音声Xに対する新HMMの確率

【数4】

$$P_{new}(X|w, \Lambda)$$

を算出する。

【数5】

$$\hat{w} = \arg \max_w [P_{new}(X|w, \Lambda)]$$

を認識結果として出力する。

【0018】係数推定部80は、新HMM確率計算部60で算出された入力音声に対する確率

【数8】

$$P_{new}(X|w, \Lambda)$$

を最大化あるいは極大化するような係数

$\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_s, \dots, \lambda_S\}$

を推定する。最適な推定式は以下のように導出することができる。初めに確率

【数9】

$$P_{new}(X|w, \Lambda)$$

を次式のように書き換える。

【数10】

$$\sum_Q f(Q) = \sum_{q_1=1}^N \sum_{q_2=1}^N \cdots \sum_{q_T=1}^N f(q_1, q_2, \dots, q_T)$$

【数14】

$$\sum_S g(S) = \sum_{s_1=1}^S \sum_{s_2=1}^S \cdots \sum_{s_T=1}^S g(s_1, s_2, \dots, s_T)$$

\*

$$Q(\Lambda, \Lambda') \geq Q(\Lambda, \Lambda) \Rightarrow P_{new}(X|w, \Lambda') \geq P_{new}(X|w, \Lambda)$$

また

※ ※ 【数16】

$$\ln[P_{new}(X, Q, S|w, \Lambda)] = \sum_{i=1}^T \left\{ \ln[a_{q_{i-1}q_i}] + \ln[\lambda'_i] + \ln[b_{q_i}^{(i)}(x_i)] \right\}$$

【数17】

$$\frac{\partial}{\partial \lambda'_s} \ln[P_{new}(X, Q, S|w, \Lambda)] = \sum_{i=1}^T \frac{\delta_{s_i}}{\lambda'_s}$$

★ここで記号 $\delta_{s_i}$ はクロネッカーのデルタ記号である。

よって

【数18】

$$\begin{aligned} \frac{\partial Q(\Lambda, \Lambda')}{\partial \lambda'_s} &= \sum_Q \sum_S P_{new}(X, Q, S|w, \Lambda) \sum_{i=1}^T \frac{\delta_{s_i}}{\lambda'_s} - \nu \\ &= \sum_{q_1=1}^N \cdots \sum_{q_T=1}^N \sum_{s_1=1}^S \cdots \sum_{s_T=1}^S \left\{ \prod_{i=1}^T a_{q_{i-1}q_i} \lambda'_i b_{q_i}^{(i)}(x_i) \right\} \left\{ \sum_{i=1}^T \frac{\delta_{s_i}}{\lambda'_s} \right\} - \nu \\ &= \frac{1}{\lambda'_s} \sum_{i=1}^T P_{new}(X, s_i = s|w, \Lambda) - \nu \end{aligned}$$

Q関数を極大化する $\Lambda'$ に対する条件式

【数19】

$$\frac{\partial Q(\Lambda, \Lambda')}{\partial \lambda'_s} = 0$$

より

【数20】

$$\lambda'_s = \frac{1}{\nu} \sum_{i=1}^T P_{new}(X, s_i = s|w, \Lambda)$$

が得られる。ここで、制約条件

【数21】

$$\sum_{s=1}^S \lambda'_s = 1$$

からラグランジェ未定係数を求めると、最終的に係数 $\lambda'_s$ の再推定式は次式になる。

【数22】

$$\lambda'_s = \frac{\sum_{i=1}^T P_{new}(X, s_i = s|w, \Lambda)}{T \cdot P_{new}(X|w, \Lambda)}$$

\* 【0019】また、記号 $\nu$ は、係数に対する制約条件を導入するためのラグランジェ未定係数である。簡単な計算により以下の関係があることがわかる。

【数15】

★ここで記号 $\delta_{s_i}$ はクロネッカーのデルタ記号である。

よって

【数18】

したがって、係数を、例えば、次のような初期値 $\lambda_s = 1/S$ から出発して、上記の再推定式を用いて逐次更新していくことにより、最適（確率を極大にする）な係数を算出することができる。係数の初期値としては各特定話者HMMによる確率値を正規化して用いてもよい。

【0020】上記の再推定式中の確率値

【数23】

$$P_{new}(X, s_i = s|w, \Lambda)$$

と

【数24】

$$P_{new}(X|w, \Lambda)$$

40

について、漸化式を用いた効率的な計算法を以下に示しておく。

【0021】前向き累積確率の時刻 $t=1$ における初期値を次式で定義する。

【数25】

$$\alpha_1(i, s) = a_{i-1} \lambda_s b_i^{(s)}(x_1)$$

時刻 $t$ における前向き累積確率は、

【数26】

$$\alpha_i(i) \equiv \sum_{s=1}^9 \alpha_i(i, s)$$

【数 2 7】

$$\alpha_i(i, s) = \sum_{j=1}^N \alpha_{i-1}(j) a_{ji} \lambda_s b_i^{(s)}(x_i)$$

で計算され、確率

【数 2 8】

$$P_{new}(X|w, \Lambda) = \sum_{i=1}^N \alpha_T(i)$$

で算出される。

【0 0 2 2】次に、後向き累積確率の時刻  $t = T$  における初期値を次式で定義する。

【数 2 9】

$$\beta_T(i) \equiv 1$$

時刻  $t$  における後向き累積確率は、

【数 3 0】

$$\beta_i(i) = \sum_{j=1}^N a_{ji} \left\{ \sum_{s=1}^9 \lambda_s b_j^{(s)}(x_{i+1}) \right\} \beta_{i+1}(j)$$

で計算され、確率

【数 3 1】

$$P_{new}(X, s_i = s | w, \Lambda) = \sum_{i=1}^N \alpha_i(i, s) \beta_i(i)$$

(6)

特開平 1 0 - 2 6 8 8 9 3

10

で算出される。

【0 0 2 3】以上により、係数推定部 8 0 において、最適な係数が効率的に計算されることが示された。推定された係数は係数記憶部 4 0 に記憶される。

【0 0 2 4】

【発明の効果】本発明の音声認識装置によれば、予め用意した複数の特定話者の中に、新しい話者に類似した話者が含まれていなかった場合にも、従来のように無理失理にあまり類似していない話者を選んでしまうことなく、特定話者 HMM の線形結合を用いることにより、新話者に最適な新 HMM を作成することができる。

【0 0 2 5】また、この特定話者 HMM の線形結合を定める係数として、HMM の枠組みにおいて最適な推定値を与えることができる。

【図面の簡単な説明】

【図 1】本発明による音声認識装置の一実施形態を示す構成ブロック図である。

【符号の説明】

- |     |               |
|-----|---------------|
| 1 0 | 音声分析部         |
| 2 0 | 特徴ベクトル出力確率計算部 |
| 3 0 | 特定話者 HMM 記憶部  |
| 4 0 | 係数記憶部         |
| 5 0 | 単語辞書記憶部       |
| 6 0 | 新 HMM 確率計算部   |
| 7 0 | 認識結果出力部       |
| 8 0 | 係数推定部         |



【図1】

